On testing for a functional relationship between mean and variance, with applications to regression.

Anila Wijesinha Cornell University 18
~~Douglas S. Robson~~ 38 ~~Cornell University~~

Transforming data can sometimes have adverse effects on a regression analysis. One such effect, the inconsistency of critical parameter estimates, may be detected by testing for a particular functional relationship between the mean and variance of the transformed dependent variable in the regression.

This paper obtains an approximate likelihood ratio statistic which tests for this particular relationship between the mean and variance, under the assumption of normality.

Significance levels of the test ~~and~~ were obtained and the properties of robustness against non-normality and power against a family of alternatives were studied (using simulation techniques) for various sample sizes.

~~Asymptotic analytic results were obtained~~ for ~~the~~ exact ~~distribution of the test~~ statistic.

ON TESTING FOR A FUNCTIONAL RELATIONSHIP BETWEEN MEAN AND VARIANCE, WITH APPLICATIONS TO REGRESSION

BU-736-M by May, 1981

Anila Wijesinha and D. S. Robson
Biometrics Unit, Cornell University, Ithaca, New York

1. INTRODUCTION

Transformation of data has long been a means of rendering variables more tractable to statistical analyses. In particular, transformations are utilized in regression analysis with numerous objectives in view, the most prevalent of which is the achievement of a convenient form of the regression equation. The mean of a non-linear transform is not the transform of the mean, however, so the transform of a correct specification of the regression function on the original scale is not a correct specification for the transformed scale. Sometimes the detrimental consequences of a non-linear transform of this nature, may outweigh the computational convenience attained by its application.

This paper discusses the possible disadvantages that arise due to the use of transformations in this context, and obtains some mathematical results whereby certain biases that occur in parameter estimation after transformation may be detected through the detection of certain structural forms of variance heterogeneity within the regression model. Considerations are here restricted to the case of variance being proportional to a specified power of the mean. An approximation to the likelihood ratio test (Neyman and Pearson, 1931) was developed for this case with the added assumption of normality, and small sample properties of the test were investigated by simulation methods.

Paper No. BU-736-M in the Biometrics Unit.

2. THE PROBLEM AND ITS MOTIVATION

We consider a regression framework where interest lies in the functional dependency of a variable Y on a vector of one or more independent variables $\underline{X}$. Three desirable properties of a statistical relationship between Y and $\underline{X}$ which facilitate the application of regression techniques are:

1. linearity in the unknown parameters, or some other convenient form of the regression equation;
2. normality of the distribution of Y conditional on $\underline{X}$;
3. homoscedasticity of Y conditional on $\underline{X}$.

The above desiderata provide computational convenience in the application of estimation procedures such as least squares, and conform with the statistical theory which underlies the development of most linear estimation procedures and tests of hypotheses, thus enabling the use of well-established methods such as t-tests, F-tests and the analysis of variance in the regression analysis.

Unfortunately, it is rarely true that all three desiderata are inherent in variables on the original scale, and often, transformations are applied in order to achieve some or all of them. The simultaneous fulfillment of all three may not be achievable, however, and untoward effects of the transformation may sometimes cancel the benefits of any partial fulfillment. A transformation to achieve linearity of regression, for example, may have such dire consequences as loss of consistency of critical parameter estimates, a loss far outweighing the computational convenience of the achieved linearity.

We consider here a situation where, on the original scale, the regression function has the correctly postulated form $\eta(\underline{X})$ with

$$E[Y|\underline{X}] = \eta(\underline{X}) ,$$

and a transformation f is applied to achieve normality and a convenient form of the deterministic function given by $f\big(\eta(\underline{X})\big)$, but which may not achieve homoscedasticity.

Transformations which are designed to achieve homoscedasticity are attempts to exploit the regularity in the heteroscedasticity ordinarily present on the untransformed scale, as manifested by a functional relationship between mean and variance on this scale. Our interest lies in the behavior of the regression after transformation, when heteroscedasticity is of such a known functional form.

Using a Taylor series expansion for $f(Y|\underline{X})$ about $\eta(\underline{X})$ we have:

$$f(Y|\underline{X}) = f\big(\eta(\underline{X})\big) + \sum_{k=1}^{\infty} \frac{\big(Y - \eta(\underline{X})\big)^k}{k!} f^{(k)}\big(\eta(\underline{X})\big)$$

where

$$f^{(k)}\big(\eta(\underline{X})\big) = \left.\frac{d^k f(t)}{dt^k}\right|_{t=\eta(\underline{X})} .$$

Therefore, taking expectations, to second order approximation of a Taylor series expansion, we have

$$\begin{aligned} E[f(Y|\underline{X})] &\doteq E\Big[f\big(\eta(\underline{X})\big)\Big] + E[Y - \eta(\underline{X})]f^{(1)}\big(\eta(\underline{X})\big) \\ &\quad + \frac{E[Y - \eta(\underline{X})]^2}{2!} f^{(2)}\big(\eta(\underline{X})\big) \\ &= f\big(\eta(\underline{X})\big) + \sigma^2_{Y|\underline{X}} \frac{f^{(2)}\big(\eta(\underline{X})\big)}{2} \qquad (2.1) \end{aligned}$$

where

$$\sigma^2_{Y|\underline{X}} = V[Y|\underline{X}] = \text{variance of } Y \text{ conditional on } \underline{X} .$$

Now, we also know that to a first order approximation of a Taylor series expansion,

$$V[f(Y|\underline{X})] \doteq \left[f^{(1)}\left(\eta(\underline{X})\right)\right]^2 V[Y|\underline{X}] \;;$$

i.e.,

$$\sigma^2_{f(Y|\underline{X})} \doteq \left[f^{(1)}\left(\eta(\underline{X})\right)\right]^2 \sigma^2_{Y|\underline{X}} \,.$$

So, on substitution for $\sigma^2_{Y|\underline{X}}$ in (2.1), we have

$$E[f(Y|\underline{X})] \doteq f\left(\eta(\underline{X})\right) + \frac{f^{(2)}\left(\eta(\underline{X})\right)}{\left[f^{(1)}\left(\eta(\underline{X})\right)\right]^2} \cdot \frac{\sigma^2_{f(Y|\underline{X})}}{2} \,. \qquad (2.2)$$

It is now seen from (2.2) that the expected value of the transformed variable is not, in general, the transformation of the expected value of the original variable, except under linear transformations where

$$f^{(k)}(t) = 0 \qquad \forall\ k \geq 2,\ t \in R \,.$$

Under all other types of transformations, to second order approximation of a Taylor series, the expectation of the transformed variable is biased by a term which is a function of the first and second derivatives of the transformation evaluated at the expectation of the original variable, and the variance of the transformed variable.

We shall refer to the term

$$E[f(Y|\underline{X})] - f[E(Y|\underline{X})] = E[f(Y|\underline{X})] - f[\eta(\underline{X})]$$

$$\doteq \frac{f^{(2)}\left(\eta(\underline{X})\right)}{\left[f^{(1)}\left(\eta(\underline{X})\right)\right]^2} \cdot \sigma^2_{f(Y|\underline{X})}$$

of (2.2) as '<u>the</u> <u>bias</u>', under our assumption of a correct model specification on the original scale.

Interest lies in the form of the bias under commonly occurring transformations. We consider, therefore, three of the more common transformations; i.e., power transformations, log transformations which are the limiting case of power transformations as the power tends to limit zero, and exponential transformations.

We now formalize the consideration of the bias under the above-mentioned three types of transformations by considering two larger classes of transformations, $\mathcal{U}_1$ and $\mathcal{U}_2$, defined in terms of their first and second derivatives as follows:

$$\mathcal{U}_1 = \left\{ f : [f^{(1)}(t)]^2 = cf^{(2)}(t),\ \forall\ t \in R \right\} ,$$

$$\mathcal{U}_2 = \left\{ f : [f^{(1)}(t)]^2 = c'f^{(2)}(t)f(t),\ \forall\ t \in R \right\} ,$$

where $R = (-\infty, \infty)$, $f^{(k)}(t) = \dfrac{d^k f(t)}{dt^k}$, $k \in \{1,2\}$ and c, c' are constants with respect to t. We note that:

$$\{\text{power transformations}\} \subset \mathcal{U}_2 ,$$

$$\{\text{log transformations}\} \subset \mathcal{U}_1$$

and

$$\{\text{exponential transformations}\} \subset \mathcal{U}_2 .$$

Thus all three families of transformations belong to $\mathcal{U}_1 \cup \mathcal{U}_2$. We note that:

$$f_1 \in \mathcal{U}_1 \Rightarrow \text{bias} = \frac{f_1^{(2)}\big(\eta(\underline{X})\big)}{\left[f_1^{(1)}\big(\eta(\underline{X})\big)\right]^2} \cdot \frac{\sigma^2_{f_1}(Y|\underline{X})}{2}$$

$$= \frac{\sigma^2_{f_1}(Y|\underline{X})}{2c}$$

$$= \sigma^2_{f_1}(Y|\underline{X}) \cdot \text{constant}$$

and

$$f_2 \in \mathcal{U}_2 \Rightarrow \text{bias} = \frac{f_2^{(2)}\big(\eta(\underline{X})\big)}{\left[f_2^{(1)}\big(\eta(\underline{X})\big)\right]^2} \cdot \frac{\sigma^2_{f_2}(Y|\underline{X})}{2}$$

$$= \frac{\sigma^2_{f_2}(Y|\underline{X})}{2c'f_2\big(\eta(\underline{X})\big)}$$

$$= \frac{\sigma^2_{f_2}(Y|\underline{X})}{f_2\big(\eta(\underline{X})\big)} \cdot \text{constant} .$$

We now restrict ourselves to heteroscedasticity of the transformed variable $f(Y|\underline{X})$ which has a functional relationship to $\eta(\underline{X})$ given by

$$\sigma^2_{f(Y|\underline{X})} \propto \left[f\big(\eta(\underline{X})\big)\right]^p$$

for a prescribed p, and examine the form which is taken by the bias.

Proposition 2.1: To second order approximation of a Taylor series expansion, for given p,

$$f \in \mathcal{U}_1 \text{ and } \sigma^2_{f(Y|\underline{X})} \propto \left[f\big(\eta(\underline{X})\big)\right]^p \Rightarrow E[f(Y|\underline{X})]$$

$$= f\big(\eta(\underline{X})\big) + \left[f\big(\eta(\underline{X})\big)\right]^p \cdot \text{constant} .$$

Proof: $E[f(Y|\underline{X})] = f\big(\eta(\underline{X})\big) + \text{bias}$

$$= f\big(\eta(\underline{X})\big) + \sigma^2_{f(Y|\underline{X})} \cdot \text{constant}$$

$$= f\big(\eta(\underline{X})\big) + \left[f\big(\eta(\underline{X})\big)\right]^p \cdot \text{constant} .$$

Corollary 2.1.1: *Transformations which achieve constant variance* ($p = 0$) *will bias the expectation by a constant term. That is,*

$$\sigma^2_{f(Y|\underline{X})} = c \Rightarrow E[f(Y|\underline{X})] = f\big(\eta(\underline{X})\big) + c' ,$$

for some constants c, c'.

Corollary 2.1.2: *Transformations which achieve constant variance and linearity of the deterministic function* ($p = 0$ *and* $f\big(\eta(\underline{X})\big) = \underline{\underline{X}}\underline{\beta}$, *where* $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $k \in N$) *will bias the intercept parameter of the linear regression. That is,*

$$\sigma^2_{f(Y|\underline{X})} = c \text{ and } f\big(\eta(\underline{X})\big) = \underline{\underline{X}}\underline{\beta} \Rightarrow E[f(Y|\underline{X})] = \underline{\underline{X}}\underline{\beta}' ,$$

where $\underline{\beta}' = (\beta_0 + c', \beta_1, \cdots, \beta_k)^T$ *for some constant* c'. *Hence,* $E(\hat{\beta}_0) = \beta_0 + c'$ *and* $E(\hat{\beta}_i) = \beta_i$, $\forall\ 1 \le i \le k$.

Corollary 2.1.3: *Transformations which achieve variance proportional to the deterministic function* ($p = 1$) *will achieve a bias that is proportional to the deterministic function. That is,*

$$\sigma^2_{f(Y|\underline{X})} \propto f\big(\eta(\underline{X})\big) \Rightarrow E[f(Y|\underline{X})] = f\big(\eta(\underline{X})\big)(1 + c) ,$$

for some constant c.

Corollary 2.1.4: *Transformations which achieve variance proportional to the deterministic function and linearity of the deterministic function* ($p = 1$ *and* $f\big(\eta(\underline{X})\big) = \underline{\underline{X}}\underline{\beta}$, *where* $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $k \in N$) *will achieve a bias that is proportional for all parameters. That is,*

$$\sigma^2_{f(Y|\underline{X})} \propto \underline{\underline{X}}\underline{\beta} \text{ and } f\big(\eta(\underline{X})\big) = \underline{\underline{X}}\underline{\beta} \Rightarrow E[f(Y|\underline{X})] = \underline{\underline{X}}\underline{\beta}' ,$$

where $\underline{\beta}' = \underline{\beta}(1 + c)$ *for some constant* c. *Thus*

$$\frac{E(\hat{\beta}_i)}{E(\hat{\beta}_j)} = \frac{\beta_i(1 + c)}{\beta_j(1 + c)} = \frac{\beta_i}{\beta_j} , \forall\ 0 \le i, j \le k .$$

Proposition 2.2: To second order approximation of a Taylor series expansion, for given p,

$$f \in \mathcal{U}_2 \text{ and } \sigma^2_{f(Y|\underline{X})} \propto \left[f\left(\eta(\underline{X})\right)\right]^p \Rightarrow E[f(Y|\underline{X})]$$
$$= f\left(\eta(\underline{X})\right) + \left[f\left(\eta(\underline{X})\right)\right]^{p-1} \cdot \text{constant} .$$

Proof: $E[f(Y|\underline{X})] = f\left(\eta(\underline{X})\right) + \text{bias}$

$$= f\left(\eta(\underline{X})\right) + \frac{\sigma^2_{f(Y|\underline{X})}}{f\left(\eta(\underline{X})\right)} \cdot \text{constant}$$

$$= f\left(\eta(\underline{X})\right) + \left[f\left(\eta(\underline{X})\right)\right]^{p-1} \cdot \text{constant} .$$

Corollary 2.2.1: Transformations which achieve variance proportional to the deterministic function (p = 1) will bias the expectation by a constant term. That is,

$$\sigma^2_{f(Y|\underline{X})} \propto f\left(\eta(\underline{X})\right) \Rightarrow E[f(Y|\underline{X})] = f\left(\eta(\underline{X})\right) + c$$

for some constant c.

Corollary 2.2.2: Transformations which achieve variance proportional to the deterministic function and linearity of the deterministic function (p = 1 and $f\left(\eta(\underline{X})\right) = \underline{\underline{X}}\underline{\beta}$, where $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $k \in N$) will bias the intercept parameter of the linear regression. That is,

$$\sigma^2_{f(Y|\underline{X})} \propto \underline{\underline{X}}\underline{\beta} \text{ and } f\left(\eta(\underline{X})\right) = \underline{\underline{X}}\underline{\beta} \Rightarrow E[f(Y|\underline{X})] = \underline{\underline{X}}\underline{\beta}'$$

where $\underline{\beta}' = (\beta_0 + c, \beta_1, \cdots, \beta_k)^T$ for some constant c. Hence $E(\hat{\beta}_0) = \beta_0 + c$ and $E(\hat{\beta}_i) = \beta_i$, $\forall\ 1 \leq i \leq k$.

<u>Corollary 2.2.3</u>: <u>Transformations which achieve standard deviation proportional to the deterministic function</u> (p = 2) <u>will achieve a bias that is proportional to the deterministic function. That is,</u>

$$\sigma^2_{f(Y|\underline{X})} \propto \left[f\left(\eta(\underline{X})\right)\right]^2 \Rightarrow E[f(Y|\underline{X})] = f\left(\eta(X)\right)(1 + c)$$

<u>for some constant</u> c.

<u>Corollary 2.2.4</u>: <u>Transformations which achieve standard deviation proportional to the deterministic function and linearity of the deterministic function</u> (p = 2 <u>and</u> $f\left(\eta(\underline{X})\right) = \underline{X}\underline{\beta}$, where $\underline{\beta} = (\beta_0, \beta_1, \cdots, \beta_k)^T$, $k \in N$) <u>will achieve a bias which is proportional for all parameters. That is,</u>

$$\sigma^2_{f(Y|\underline{X})} \propto (\underline{X}\underline{\beta})^2 \text{ and } f\left(\eta(\underline{X})\right) = \underline{X}\underline{\beta} \Rightarrow E[f(Y|\underline{X})] = \underline{X}\underline{\beta}'$$

<u>where</u> $\underline{\beta}' = \underline{\beta}(1 + c)$ <u>for some constant</u> c. <u>Thus</u>

$$\frac{E(\hat{\beta}_i)}{E(\hat{\beta}_j)} = \frac{\beta_i}{\beta_j}, \ \forall\ 0 \le i,\ j \le k .$$

In summary, for $f \in \mathcal{U}_1 \cup \mathcal{U}_2$,

$$f \in \mathcal{U}_1 \text{ and } \sigma^2_{f(Y|\underline{X})} = \text{constant} \Rightarrow \text{constant bias} .$$

$$f \in \mathcal{U}_2 \text{ and } \sigma^2_{f(Y|\underline{X})} \propto f\left(\eta(\underline{X})\right) \Rightarrow \text{constant bias} .$$

$$f \in \mathcal{U}_1 \text{ and } \sigma^2_{f(Y|\underline{X})} \propto f\left(\eta(\underline{X})\right) \Rightarrow \text{bias} \propto f\left(\eta(\underline{X})\right) .$$

$$f \in \mathcal{U}_2 \text{ and } \sigma^2_{f(Y|\underline{X})} \propto \left[f\left(\eta(\underline{X})\right)\right]^2 \Rightarrow \text{bias} \propto f\left(\eta(\underline{X})\right) .$$

In view of the two propositions and their corresponding corollaries, it is seen that for transformations in the class $\mathcal{U}_1 \cup \mathcal{U}_2$, this type of variance structure of the transformed variable, in relation to the transformed expectation, determines the bias that occurs in the estimation of the original parameters, after transformation. Thus, if this functional relationship between the variance and the deterministic function could be detected, then the type of bias that could occur in the circumstances would be known, and measures could be taken to either adjust for it in parameter estimation, or incorporate the information obtained about the bias in subsequent tests of hypotheses involving the parameters.

Thus, this motivates the need for a test statistic of the null hypothesis

$$H_0 : \sigma^2_{f(Y|\underline{X})} \propto \left[f\left(\eta(\underline{X})\right)\right]^p$$

versus the alternative

$$H_A : \sigma^2_{f(Y|\underline{X})} \not\propto \left[f\left(\eta(\underline{X})\right)\right]^p$$

for any given prescribed p.

This paper concerns itself with the development of such a test statistic, and the evaluation of its properties. In order to develop this test procedure, the concept of the generalized likelihood ratio criterion is utilized.

3. DEVELOPMENT OF A GENERALIZED LIKELIHOOD RATIO STATISTIC

3.1. Derivation of the Exact Generalized Likelihood Ratio Statistic

We consider a situation where k independent observations (replicates) are available from each of n normal populations:

$$Z_{ij} = \mu_i + \epsilon_{ij}, \quad \forall\ 1 \le i \le n,\ 1 \le j \le k .$$

In order to develop the likelihood ratio, we consider the most general form of the $\{\mu_i\}$ which is given by

$$\mu_i = \mu_i(\underline{\theta}), \quad \forall\ 1 \le i \le n ,$$

where $\underline{\theta}$ is a vector of unknown independent parameters defining any underlying dependence structure among the μ_i's. If the dimension of $\underline{\theta}$ is n or greater, then the introduction of this parameterization of the means would be redundant.

Under the assumption of normality, the distribution of the Z_{ij}'s is entirely specified by $\underset{\sim}{\theta}$, which defines the means $\{\mu_i\}$, and by the parameters which define the variances $\{\sigma_i^2\}$. The likelihood function for the n X k independent observations is given by

$$L = (2\pi)^{-\frac{nk}{2}} \exp\left\{-\tfrac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{\left(Z_{ij}-\mu_i(\underset{\sim}{\theta})\right)^2}{\sigma_i^2}\right\}\prod_{i=1}^{n}(\sigma_i^2)^{-\frac{k}{2}} . \tag{3.1}$$

A generalized likelihood ratio statistic for testing the hypothesis

$$H_0 : \sigma_i^2 = [\mu_i(\underset{\sim}{\theta})]^p c^2, \quad \forall\ 1 \le i \le n$$

versus the alternative

$$H_A : \sigma_i^2 \neq [\mu_i(\underset{\sim}{\theta})]^p c^2$$

for a given p, is obtained by maximizing L in Ω_0, the parameter space under H_0, and in $\Omega_0 \cup \Omega_A$, the entire parameter space, and then taking the ratio (λ) of the two maxima (Kendall and Stuart, 1972); thus,

$$\lambda = \frac{L_{max}(\Omega_0)}{L_{max}(\Omega_0 \cup \Omega_A)} .$$

On solving the maximum likelihood equations under Ω_0 and $\Omega_0 \cup \Omega_A$, this reduces to

$$\lambda = \frac{\prod_{i=1}^{n}\left(\sum_{j=1}^{k}\frac{(Z_{ij}-\tilde{\mu}_i)^2}{k}\right)^{\frac{k}{2}}}{\left\{\frac{1}{nk}\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{(Z_{ij}-\hat{\mu}_i)^2}{\hat{\mu}_i^p}\right\}^{\frac{nk}{2}}\prod_{i=1}^{n}\hat{\mu}_i^{\frac{pk}{2}}} ,$$

where $\forall\ 1 \le i \le n$, $\tilde{\mu}_i$ is the maximum likelihood estimate of μ_i under $\Omega_0 \cup \Omega_A$ and $\hat{\mu}_i$ is the maximum likelihood estimate of μ_i under Ω_0.

The H_0-distribution of $-2\log\lambda$ has been proved to be asymptotically chi-squared (Kendall and Stuart, 1972) with degrees of freedom given by the difference in dimension of the parameter spaces Ω_0 and $\Omega_0 \cup \Omega_A$, which in this case is $n-1$, since under Ω_0, $\underline{\theta}$ and c^2 define the set of distributions, while under $\Omega_0 \cup \Omega_A$, $\underline{\theta}$ and the n σ_i^2 define the system. Thus, the statistic λ provides a test of the hypothesis of interest for general p.

Unfortunately, the maximum likelihood equations are non-trivial, even in the simplest case where $\underline{\theta} = (\mu_1, \cdots, \mu_n)$, and in general a closed form solution for the parameters does not exist. This implies that iterative techniques are required in order to solve the equations. In view of these difficulties, the application of the likelihood ratio test is no longer simple, which undermines its usefulness as a convenient tool. This suggests the need for a simpler, more easily calculated statistic which could test the same hypothesis H_0. One such statistic is developed in the next section.

3.2. Development of an Approximate Generalized Likelihood Ratio Test Statistic

The difficulty in calculating λ arose from difficulties in solving the maximum likelihood equations for the estimates of the parameters of the model. Therefore the possibility of replacing the maximum likelihood estimates $\{\tilde{\mu}_i\}$ and $\{\hat{\mu}_i\}$ by readily available estimates of the μ_i's and thereby creating a modified version of λ, immediately comes to mind. Since replicate observations are available from each of the n populations, we consider an approximate statistic λ_a which is obtained by merely replacing the estimates $\tilde{\mu}_i$ and $\hat{\mu}_i$ of each μ_i, by the sample means $\bar{Z}_{i.}$ of the k sample values from each population, which are also consistent estimates of the μ_i's. That is,

$$\lambda_a = \frac{\prod_{i=1}^{n}\left(\sum_{j=1}^{k}(Z_{ij} - \bar{Z}_{i.})^2\right)^{\frac{k}{2}}}{\prod_{i=1}^{n}(\bar{Z}^p_{i.})^{\frac{k}{2}}\left\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{k}\frac{(Z_{ij} - \bar{Z}_{i.})^2}{\bar{Z}^p_{i.}}\right\}^{\frac{nk}{2}}},$$

where $\forall\ 1 \leq i \leq n$, $\bar{Z}_{i.} = \frac{1}{k}\sum_{j=1}^{k} Z_{ij}$. We may rewrite this as

$$\lambda_a = \left[\frac{\prod_{i=1}^{n}\frac{S^2_{Z_i}}{\bar{Z}^p_{i.}}}{\left(\frac{1}{n}\sum_{i=1}^{n}\frac{S^2_{Z_i}}{\bar{Z}^p_{i.}}\right)^n}\right]^{\frac{k}{2}}$$

where $\forall\ 1 \leq i \leq n$, $S^2_{Z_i} = \sum_{j=1}^{k}(Z_{ij} - \bar{Z}_{i.})^2$ is the sample sum of squares of the i^{th} population. We note that λ_a is the ratio of the geometric mean to the arithmetic mean of the n quantities $\frac{S^2_{Z_i}}{\bar{Z}^p_i}$, raised to the $\frac{nk}{2}^{th}$ power. This is a very simple statistic to calculate. It remains to be verified that the sampling distribution of λ_a or a simple function of it is operationally tractable, and that it has desirable properties as a test statistic.

We note that λ_a is, in some sense, a simple perturbation of λ. Therefore it is to be expected that the distribution of $-2\log\lambda_a$ with a slight modification would be very similar to the distribution of $-2\log\lambda$, which has already been proved to be asymptotically chi-square with $(n-1)$ degrees of freedom. In view of this idea, we approach the task of deriving the distribution of $-2\log\lambda_a$ under the null hypothesis, by initially deriving its first two moments and comparing them with that of a chi-square distribution with $(n-1)$ degrees of freedom.

We have that

$$-2\log\lambda_a = k\left\{n\log\left(\frac{1}{n}\sum_{i=1}^{n}\frac{S^2_{Z_i}}{\bar{Z}^p_i}\right) - \sum_{i=1}^{n}\frac{S^2_{Z_i}}{\bar{Z}^p_i}\right\}.$$

By expanding the above as the difference of two Taylor series, it can be shown that

$$E(-2\log\lambda_a) \doteq (n-1)\frac{k}{(k-1)}\left\{1 + \frac{1}{3(k-1)}\left(1+\frac{1}{n}\right) + \frac{p^2c^2}{2n}\sum_{i=1}^{n}\mu_i^{p-2}\left(1 - \frac{1}{(k-1)}\left(1+\frac{2}{n}\right)\right)\right\}$$

and

$$V(-2\log\lambda_a) \doteq 2(n-1)\frac{k^2}{(k-1)^2}\left\{1 + \frac{2}{3(k-1)}\left(1+\frac{1}{n}\right) + \frac{p^2c^2}{n}\sum_{i=1}^{n}\mu_i^{p-2}\left(1 - \frac{1}{(k-1)}\left(1+\frac{2}{n}\right)\right)\right\}$$

(Wijesinha, 1981). Thus, if we consider a bias adjustment given by

$$B = \frac{(k-1)}{k}\left\{1 + \frac{1}{3(k-1)}\left(1+\frac{1}{n}\right) + \frac{p^2c^2}{2n}\sum_{i=1}^{n}\mu_i^{p-2}\left(1 - \frac{1}{(k-1)}\left(1+\frac{2}{n}\right)\right)\right\}$$

we have

$$E\left[\frac{-2\log\lambda_a}{B}\right] \doteq (n-1)$$

and

$$V\left[\frac{-2\log\lambda_a}{B}\right] \doteq 2(n-1)\left(1 - \frac{p^2c^2}{3n(k-1)}\sum_{i=1}^{n}\mu_i^{p-2}\left(1+\frac{1}{n}\right)\right)$$

$$\rightarrow 2(n-1) \quad \text{for large } k .$$

For large sample sizes $\frac{-2\log\lambda_a}{B}$ thus has the first two moments of a chi-square distribution. Utilizing the same estimates of c^2 and the μ_i's that replaced the original maximum likelihood estimates in the statistic λ to obtain λ_a, a consistent estimate of B is obtained which is given by

$$\hat{B} = 1 + \frac{1}{3(k-1)}\left(1 + \frac{1}{n}\right) + \left(\frac{p^2\tilde{c}^2}{2n(k-1)} \sum_{i=1}^{n} \bar{Z}_{i.}^{p}\right)\left(1 - \frac{1}{(k-1)}\left(1 + \frac{2}{n}\right)\right).$$

The quantity $\dfrac{-2 \log \lambda_a}{\hat{B}}$ is proposed as a test statistic with $\dfrac{-2 \log \lambda_a}{\hat{B}} \overset{\text{approx.}}{\sim} \chi^2(n-1)$ for reasonably large k.

When p = 0, the statistic $\dfrac{-2 \log \lambda_a}{\hat{B}}$ reduces to Bartlett's statistic (Bartlett, 1937) for testing the homogeneity of variances for the case of equal numbers of observations from each population.

4. SMALL-SAMPLE PROPERTIES OF THE APPROXIMATE TEST STATISTIC – A MONTE CARLO STUDY

4.1. Distribution Under the Null Hypothesis

The results of Section 3.2 indicated that $\dfrac{-2 \log \lambda_a}{\hat{B}}$ had useful large-sample properties as a test statistic. Its relationship to Bartlett's statistic for testing the homogeneity of variance, and its first two moments provided justification for assuming its asymptotic distribution to be chi-square. However, in terms of general applicability, its small-sample properties are of more importance. In order to evaluate its distribution under small sample sizes, a Monte Carlo study was undertaken.

In view of Propositions 2.1, 2.2 and their associated corollaries, the simulation study concentrated on the cases p = 1 and p = 2 for the parameter of power-proportionality. The underlying n populations were taken at n equally spaced values of the independent variable in a simple linear regression. $\dfrac{-2 \log \lambda_a}{\hat{B}}$ was calculated for each sample, and its frequency distribution and sample moments tabulated for different parameter values of p, c, n and k (Table 4.1).

Table 4.1: Sample Moments and Frequency Distribution of $\frac{-2 \log \lambda_a}{\hat{B}}$ Under the Null Hypothesis (based on 500 samples).

p	c	Nominal statistics, and intervals[1] associated with nominal percentages	n 3	4		6		8	
			k 4	3	6	2	7	4	8
1	.05	Mean[2]	2.1	2.9	2.9	4.7	5.1	6.8	7.0
		Variance[3]	4.3	6.2	6.2	7.6	10.0	12.9	13.9
		0.5 ± 0.6	0.6	0.8	0.6	0.0	0.2	0.6	0.6
		1.0 ± 0.9	1.4	0.8	0.6	0.2	0.4	1.0	0.8
		5.0 ± 1.9	5.2	5.6	5.6	3.2	6.2	4.4	4.8
		10.0 ± 2.6	11.0	9.6	11.2	8.6	10.6	9.4	9.4
		25.0 ± 3.8	29.8	24.8	24.0	22.6	26.4	21.8	24.8
	.20	Mean	2.1	2.9	2.9	4.7	5.1	6.8	7.0
		Variance	4.3	6.2	6.2	7.6	10.0	13.0	14.0
		0.5 ± 0.6	0.6	0.8	0.6	0.0	0.2	0.6	0.6
		1.0 ± 0.9	1.0	0.8	0.6	0.2	0.6	1.0	0.8
		5.0 ± 1.9	5.4	5.2	5.4	3.0	5.6	4.2	4.6
		10.0 ± 2.6	10.8	10.2	11.4	8.6	10.8	9.4	9.4
		25.0 ± 3.8	29.2	25.4	23.4	21.6	26.8	22.2	25.0
2	.05		k 4	2	8	4	6	3	7
		Mean	2.2	3.0	3.0	4.9	5.0	6.9	7.0
		Variance	4.4	4.2	5.8	9.3	9.6	14.0	11.8
		0.5 ± 0.6	0.2	0.0	0.2	0.2	0.2	0.6	0.2
		1.0 ± 0.9	1.2	0.0	0.6	0.4	0.8	1.0	0.4
		5.0 ± 1.9	7.0	2.6	5.4	5.4	5.8	4.4	3.0
		10.0 ± 2.6	11.2	9.0	11.0	9.8	10.4	9.6	10.0
		25.0 ± 3.8	29.8	24.4	24.0	23.8	23.4	24.6	27.4
	.20	Mean	2.2	3.1	3.0	4.9	5.0	7.1	7.0
		Variance	4.5	4.5	6.0	9.3	9.6	14.2	11.8
		0.5 ± 0.6	0.2	0.0	0.2	0.2	0.2	0.4	0.2
		1.0 ± 0.9	1.2	0.0	0.8	0.8	0.6	1.0	0.8
		5.0 ± 1.9	6.6	3.4	5.2	4.8	6.0	4.2	1.0
		10.0 ± 2.6	12.2	9.6	10.8	10.4	11.0	10.2	9.6
		25.0 ± 3.8	31.2	25.4	24.8	24.4	23.8	25.8	26.2

1/ intervals given by $\alpha \pm 1.96\sqrt{\frac{\alpha(100-\alpha)}{500}}$; 2/ mean = n-1; 3/ variance = 2(n-1).
for α = .5, 1.0, 5.0, 10.0, 25.0.

It is seen that close overall agreement with the relevant chi-square distribution was achieved even for very small sample sizes given by $n = 3$ and $k = 3$. However, for the minimum value of $k = 2$, the results were somewhat unreliable and erratic. The results of this study confirmed that the approximate statistic was well-behaved even for small sample sizes, when the null hypothesis was true and the underlying assumption of normality was satisfied.

4.2. Power of the Test for a Class of Alternatives

A further Monte Carlo study was undertaken (Wijesinha, 1981) to evaluate the power of the approximate test against a class of relevant alternative hypotheses. The power for a test of size $\alpha = .05$ was calculated for $p = 0$, 1 and 2 when the true relationship between the mean and variance was given by

$$H_A : \sigma_i^2 = c^2 \mu_i^{p_A} \quad \forall\ 1 \leq i \leq n ,$$

while under the null hypothesis the relationship was given by

$$H_0 : \sigma_i^2 = c^2 \mu_i^{p} \quad \forall\ 1 \leq i \leq n ,$$

where $p_A \neq p$. Empirical evidence from the Monte Carlo study indicated the test to be both consistent and unbiased for this class of alternatives. Although the test achieved high power when the 'distance' between the null and alternative was large, i.e., $|p_A - p| \geq 2$, the power achieved was very low for $|p_A - p| = 1$. However, Bartlett's test, which was given by $p = 0$, performed in a similar manner to the cases $p = 1$ and 2 in all respects. The effect of sample sizes n and k on the increase of power achieved, was very clear, especially for large values of $|p_A - p|$.

4.3. Robustness Against Non-normality

A Monte Carlo study was also undertaken to assess the performance of the test under different types of skewness and kurtosis in the underlying distribution

(Wijesinha, 1981). Different types of long-tailed, short-tailed and truncated distributions replaced the normal distribution, in order to evaluate robustness of the test against non-normality. The sample frequency distribution was obtained when the underlying distribution was taken from the following:

1. t distributions with different degrees of freedom;
2. lognormal distributions with different parameters;
3. convolutions of different numbers of uniform distributions.

In all cases the underlying variables were standardized to have the same mean and constant of proportionality c^2 between the variance and the p^{th} power of the mean, in order that direct comparisons could be made among them.

The chi-square approximation to the distribution of $\frac{-2 \log \lambda_a}{\hat{B}}$ was poor for t distributions with degrees of freedom less than 10. However, for larger degrees of freedom the chi-square approximation compared very well.

The chi-square approximation was poor for lognormal distributions with a high degree of skewness, but as the skewness decreased a satisfactory approximation to the relevant chi-square was obtained for the distribution of $\frac{-2 \log \lambda_a}{\hat{B}}$. This indicated that the statistic was robust only against low degrees of non-normal skewness, and also truncation.

Convolutions of one and two uniform distributions fared very poorly in the chi-square approximation for the distribution of $\frac{-2 \log \lambda_a}{\hat{B}}$. Values of the upper tail were very scarce. However, convolutions of three or more uniform distributions conformed well to the expected chi-square distributions.

In view of the results of the Monte Carlo study, it appeared that the test statistic was only robust against moderate degrees of non-normal skewness and kurtosis, truncation and heavy-tailed distributions.

5. SUMMARY

Certain detrimental effects of the use of transformations have been emphasized, and a method is outlined in this paper, of detecting these untoward effects by means of a statistical test. The test that is developed has been shown to have important and useful properties. A Monte Carlo study served to fortify its usefulness by providing evidence of desirable small-sample properties of the test.

The test is proposed as a convenient tool in a preliminary investigation of data, primarily in a regression framework after a non-linear transformation has been applied to the data.

REFERENCES

Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proceedings of the Royal Statistical Society of London, Series A, 160: 268-282.

Kendall, M. G. and Stuart, A. (1972). The Advanced Theory of Statistics, Volume 2, Chapter 24, Third Edition, Griffin and Co. Ltd., London.

Neyman, J. and Pearson, E. S. (1931). On the problem of k samples. Bulletin de l'Academie Polonaise des Sciences et des Lettres, Série A, Science Mathematiques, 460-481.

Wijesinha, A. (1981). On testing for a functional relationship between mean and variance, with applications to regression. Ph.D. Thesis, Cornell University, Ithaca, New York.